

SRI's Experience with the ATIS Evaluation

Robert Moore, Douglas Appelt, John Bear,
Mary Dalrymple, and Douglas Moran

SRI International
333 Ravenswood Ave.
Menlo Park, CA 94025

Abstract

SRI International participated in the June 1990 Air Travel Information System (ATIS) natural-language evaluation. This report briefly describes the system that SRI used in the evaluation, analyzes SRI's results, and makes some recommendations for changes in the database structure and data collection system to be used for future ATIS evaluations.

The SRI ATIS System

The natural-language processing system used by SRI in the June 1990 ATIS evaluation is a derivative of the Core Language Engine (CLE) developed at SRI's Cambridge Research Centre in Cambridge, England [1]. At present, the main processing components of SRI's ATIS system are taken from the CLE, while the grammar, semantic interpretation rules, and lexicon are substantially new. The system divides query processing into the following phases:

- Lexical lookup
- Syntactic parsing
- Semantic interpretation and selectional filtering
- Quantifier scoping
- Database query generation
- Query optimization
- Database retrieval

The syntactic and semantic rules used in the parsing and interpretation phases are expressed in a unification-based formalism. The parser is based on a left-corner parsing algorithm for context-free grammar that has been generalized to apply to unification grammar by substituting unification for identity checks in dealing with grammatical category expressions. An attribute/value notation for feature constraints is provided for the grammar writer, but this notation is compiled into ordinary term structures by assigning, for each major category symbol, an argument position for each feature that can occur with that category. Grammatical unification is then implemented simply as term unification in Prolog, which is the implementation language used in the system.

In the semantic interpretation phase, logical form expressions are computed bottom-up by applying semantic interpretation rules keyed to the syntax rules. Terms in the logical form language have semantic sorts associated with them, and functors are restricted with respect to the sorts of their arguments. These sort restrictions are applied as the logical forms are constructed, acting as a filter on the structures produced by the syntactic and semantic rules. The outputs of the semantic interpretation phase are quasi-logical forms in which the scope of quantified noun phrases has not yet been determined. Quantifier scope is assigned in the next phase of processing.

At this point in processing, a database-independent formal representation of the meaning of the query has been assigned. This is transformed into a database query, principally by replacing the logical-form constants and predicates derived from the lexicon with database predicates and constants. The query is then re-ordered, if necessary, to optimize database retrieval, and the answer is retrieved from the database, which is stored as a set of Prolog clauses.

Analysis of Results

In the blind test conducted for the June 1990 ATIS evaluation, out of 90 test queries, the SRI system produced correct answers for 25, incorrect answers for 5, and no answer for 60. Thus, the dominant factor in the performance of the SRI system was that most queries failed to get through all stages of processing. Table 1 displays the number and percentage of the queries that failed to get past various levels of processing.

These numbers should be regarded at best as only an approximation of the performance of the different components of the system, for two reasons. First, no attempt has been made to judge the correctness of the output of individual system phases, only to determine whether the phase produced an answer at all. Second, the failure rate of the later phases of processing would probably have been higher if more queries had gotten past the earlier phases of processing.

With these caveats, the results seem to indicate that most of the difficulties arose in the semantic interpretation phase and the database query generation phase. The grammar seemed to provide fairly good coverage of the syntactic constructions used, and the lexicon performed surprisingly well given that the vocabulary in the test was completely uncontrolled. Undoubtedly, many of the parsing and interpretation failures

Level	Number	Percent
Lexicon	1	1.1
Parsing	14	15.5
Interpretation	28	31.1
DB query gen.	17	18.9

Table 1: Analysis of SRI ATIS Results

were due to the absence of some of the necessary lexical entries for particular words, but almost no words in the test material were totally absent from the lexicon.

The semantic rules and the database query generator are, in fact, the parts of the system that are the most recent in origin and must be regarded as far from complete, independently of how they performed on this evaluation. Our main conclusion, then, is simply that much more work is needed on these parts of the system.

Recommendations

In the course of working with the ATIS database and development data, it seemed to the SRI team that there are a number of changes in the database structure and the data collection system that would result in more interesting data being collected, and that would make system development easier for ATIS system builders. The philosophy that Texas Instruments followed in setting up the data collection system was to present information to the subject in a way that mirrored as closely as possible the way the information is presented in the printed Official Airline Guide (OAG). We believe that an attempt should be made to tailor the presentation of information to the capabilities of eventual interactive spoken-language computer systems rather than the printed page. The current ATIS data collection system presents a lot of information to the subject in response to most queries, but does so by using many abbreviated codes and column headings that are compressed in order to fit as much information as possible on one line of the screen. This is appropriate for a printed document, because of the difficulties of cross-referencing multiple tables in different parts of a printed volume, and because of the need to keep the physical size of the volume down to manageable proportions. Neither of these reasons applies to an interactive spoken language computer system where cross-referencing is easily performed by the system, and much larger volumes of data are easily handled.

We would recommend that the data collection system be modified to present less information in response to most queries, but to present that information in a fuller, less abbreviated form. It has been widely noted that about one-third of the ATIS queries collected so far are about the meaning of codes or abbreviated column headings in the displays, rather than about the domain. If fewer columns were presented in each display, it would be possible to avoid the use of many of these abbreviations. Moreover, it might prompt subjects to ask more follow-up questions to retrieve the information not displayed, generating a wider range of queries in the domain of air travel planning.

Implementing this recommendation will require changing not only the displays, but also the structure of the database, so that database tuples that differ only in information not displayed to the subject can be eliminated. Otherwise, the subject would see what appear to be duplicate answers in the display.

A number of other changes to the structure of the database would also seem to be desirable. One significant problem is the status of connecting flights. We believe it is important to devote some thought and attention to restructuring the database to put connecting flights on an equal footing with direct flights in the ATIS database. Currently, these are not even listed in the flight table, so that requests for all flights that meet certain constraints result in information only about direct flights. As a result there are almost no queries about connecting flights in the ATIS data, perhaps because the subjects are not aware of their existence. A related issue is that there is no fare information on connecting flights, because it is not presented in the printed OAG. We believe that if fare information for connecting flights cannot be obtained from OAG, then reasonable fares should be computed for them.

These seem to us to be the most important database and data collection system issues that need to be addressed for future ATIS evaluations, but there are many other smaller issues as well. We therefore suggest that a task force should be created to address these issues and decide on changes to be implemented for future ATIS evaluations.

References

- [1] Alshawi, et al, *Interim Report on the SRI Core Language Engine*, Technical Report CCSRC-5, SRI International, Cambridge Research Centre, Cambridge, England, 1988.